

Black Sheep or Scapegoats? Implementable Monitoring Policies under Unobservable Levels of Misbehavior

Berno Buechel and Gerd Muehlheusser

ABSTRACT

An authority delegates a monitoring task to an agent. Thereby, it can observe the number of detected offenders but not the monitoring intensity chosen by the agent or the resulting level of misbehavior. We provide a necessary and sufficient condition for the implementability of monitoring policies. When several monitoring intensities lead to an observationally identical outcome, only the minimum of these is implementable, which can lead to underenforcement. A comparative-statics analysis reveals that increasing the punishment can undermine deterrence, since the maximal implementable monitoring intensity decreases. When the agent is strongly intrinsically motivated to curb crime, our results are mirrored, and only high monitoring intensities are implementable. Then, higher monetary rewards for detections lead to a lower monitoring intensity and to a higher level of misbehavior.

1. INTRODUCTION

In many contexts of delegated monitoring, looking only at detection statistics need not be informative about the quality of monitoring. For example, suppose a division head of a large company reports to corporate headquarters a low number of violations against some corporate code of conduct for his division (for example, compliance with certain ethical

BERNO BUECHEL is a Postdoctoral Fellow at the Institute of Economics, University of St. Gallen, and at the Liechtenstein Institute. GERD MUEHLHEUSSER is Professor of Economics at the University of Hamburg. We are grateful to William H. J. Hubbard and an anonymous referee whose comments were very helpful in improving the paper. We also thank Eberhard Feess, Christos Litsios, Grischa Perino, Nicolas Sahuguet, Jan Schmitz, Urs Schweizer, Avi Tabbach, and Niklas Wallmeier as well as participants of the American Law and Economics Association meetings in 2014 (Chicago) and 2015 (New York) and seminars audiences at the University of Adelaide, the University of Bonn, Tel Aviv University, the University of Paderborn, and Paris West University Nanterre for their comments and suggestions.

[*Journal of Legal Studies*, vol. 45 (June 2016)]

© 2016 by The University of Chicago. All rights reserved. 0047-2530/2016/4502-0011\$10.00

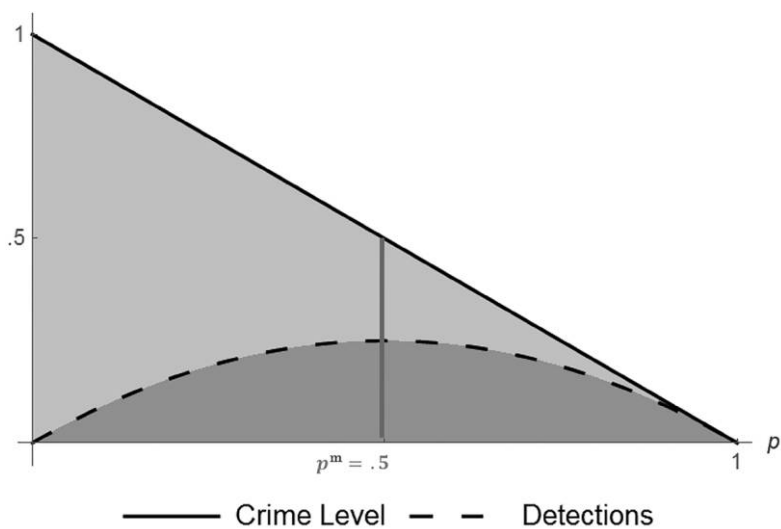
or safety standards). Then, it is not obvious what this information reveals about the true level of misbehavior in the division. A low number of detections could result from a strict monitoring policy that leads to few offenders, most of which are detected (black sheep). Instead, the monitoring policy could be lax, which leads to a large number of offenders, out of which only a few (scapegoats) are discovered. As a further example, in sports competitions it is hard for outsiders to judge what a given number of detected dopers reveals about the seriousness and intensity of antidoping measures by the respective agencies and the virulence of doping among athletes. Further examples include offenses such as tax evasion, parking violations, prostitution, trafficking, and drug dealing, in which the number of detected offenders might not be very informative about the prevalence of an illegal activity.

The common feature of these examples is that an authority delegates the task of monitoring a population of individuals to an agent. Thereby, it is an outsider in the sense that it can neither observe the monitoring intensity chosen by the agent nor the resulting level of misbehavior.¹ In contrast, the potential offenders have a good assessment of the probability of being detected, which is a standard assumption in the economic literature on enforcement (see, for example, Polinsky and Shavell 2007).² In this paper, we develop a simple model that captures the interaction among the authority, the monitoring agent, and potential offenders and that builds on the previous literature on private law enforcement with a monopolistic enforcer (see, for example, Becker and Stigler 1974; Landes and Posner 1975; Polinsky 1980; Besanko and Spulber 1989; Garoupa and Klerman 2002; Coşgel, Etkes, and Miceli 2011).

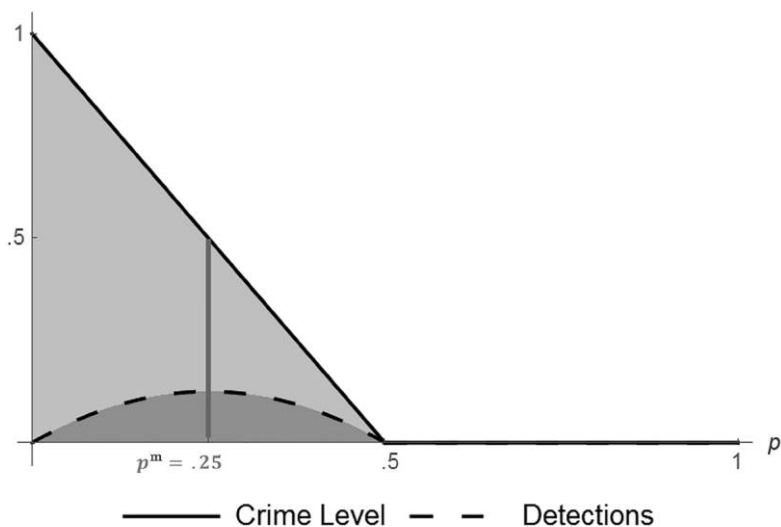
Some key features of our model can be foreshadowed with the following simple example. The enforcement effort of the monitoring agent (the inspector) translates into p , the probability that an offense is detected and punished. The number of offenses committed is decreasing in the expected punishment and, hence, decreasing in p for a given fine (T). For

1. The feature that several monitoring intensities lead to the same number of detections also applies to many other settings such as education or loan audits. However, as discussed in Section 8, it is less clear in these contexts that the authority can be considered an outsider that has to rely on the number of reported detections only.

2. A similar informational structure is also considered in the model of Arlen (1994) in the context of corporate criminal liability. Moreover, as for the case of street prostitution, regular market participants might (correctly) perceive the actual threat of being arrested by the police (let alone convicted) to be much smaller than might be presumed by outsiders (see, for example, Levitt and Venkatesh 2007).



(A) Low Fine ($T = 1$)



(B) High Fine ($T = 2$)

Figure 1. Crime level and number of detections as a function of detection probability p for two fine levels.

instance, Figure 1 depicts the number of offenses as a linear function $F(p, T) = 1 - pT$ for fine levels $T = 1$ and $T = 2$.³ Figure 1 also depicts the number of detections $pF(p, T)$ for each case (dashed curves).

The example is helpful in illustrating the main contributions of our paper compared with the existing strands of literature on law enforcement and on delegated monitoring. First, we strengthen the argument already provided in Polinsky (1980) that high monitoring intensities might not be implementable because enforcers anticipate that strong deterrence reduces revenues. In particular, we provide a full characterization of implementable monitoring policies. Thus, we do not confine attention only to a particular monitoring policy that, for instance, maximizes social welfare (see, for example, Polinsky 1980) or some other objective function of the authority (Garoupa and Klerman 2002). Intuitively, in Figure 1, for every monitoring intensity that corresponds to the decreasing part of the detection function there exists another one on the increasing part that gives rise to the same number of detected offenses and is hence observationally identical for an outside observer. However, the inspector prefers the lower of the monitoring intensities, as it involves less enforcement effort (thereby leading to more offenses), so that the largest monitoring intensity that can be implemented is p^m , where the number of detections is maximum. More generally, when several monitoring intensities give rise to the same number of detected offenses, then the agent can be induced to choose only the minimum of these.⁴ For this reason, under quite general conditions (for example, with respect to the underlying distribution of individuals' gains from the offense or the agent's effort cost function), a large set of monitoring policies cannot be implemented by the authority, even if it has unlimited funds to reward the inspector. This puts a lower bound on the number of offenses that can be achieved in this setup and can result in underenforcement compared with an efficiency benchmark.

Second, to evaluate the scope of this result, we perform a comparative-statics analysis with respect to both the distribution of gains from crime (where the previous literature has usually confined attention to the uniform case only) and the fine for detected offenders. Our results point to a novel trade-off between the severity of the punishment and the set of

3. As will become clear, linear crime functions arise when the distribution of the gains from the crime is uniform.

4. In Ichino and Muehlheusser (2008), a low monitoring intensity results as an optimal choice of the inspector, as this allows him to elicit private information from (potential) offenders.

implementable monitoring intensities: on the one hand, when the punishment is relatively strong (for example, because there are not many individuals with sufficiently large gains from the offense or because the fine is high), then the crime level tends to be low for any given monitoring intensity p . But on the other hand, the set of implementable monitoring intensities is small. In Figure 1 this trade-off can be seen by comparing the two graphs. The higher fine results in a lower number of offenses for any given monitoring intensity, but it also reduces the maximal implementable monitoring policy. In this example, the two effects just offset each other so that the minimal number of offenses $F(p^m(T), T)$ is the same for both fine levels (and equal to .5), but we also show that the latter effect may dominate, so that deterrence may in fact decrease as the punishment becomes more severe.

This result is in contrast to the standard approach in the enforcement literature in the tradition of Becker (1968), where the two components of expected punishment—the probability of detection and the fine—can be set independently from each other. As a consequence, higher fines typically lead to more deterrence.⁵ Moreover, the literature provides various reasons against Becker's stark conclusion that fines should be set as large as possible. Examples include offenders who are risk averse or heterogeneous with respect to their wealth, offenders who engage in socially undesirable avoidance activities, costs of fine collection, or the requirement that the punishment should reflect the severity of the offense (for a detailed discussion, see Polinsky and Shavell 2007). But in these frameworks, higher fines would also always lead to more deterrence. In contrast, our result suggests that in the context of delegated monitoring, even in the absence of all of these countervailing factors, optimal fines might not be too large because of the potentially detrimental effect on deterrence.⁶ Importantly, the potentially inverse relationship between the severity of punishment and deterrence is not driven by behavioral biases or irrationality, for example, on the side of the offenders. Finally, the novel trade-off identified in our framework might also add to the difficulty of the empirical literature in providing robust evidence in favor of Becker's

5. One exception is the model of Nussim and Tabbach (2009), in which in addition to the decision about their level of criminal behavior, offenders can engage in avoidance activities. Further exceptions include frameworks of juror behavior (Andreoni 1991; Feess and Wohlschlegel 2009), inspection games (Tsebelis 1990), and corruption (Kugler, Verdier, and Zenou 2005).

6. Note that this issue cannot be mitigated by replacing fines with imprisonment because our argument holds for any type of punishment.

deterrence hypothesis, apart from the well-known methodological issues (see, for example, Levitt 1997; Di Tella and Schargrodsky 2004; Levitt and Miles 2007).

Third, we consider an extension of the model in which we allow the inspector to be crime sensitive in the sense that she directly benefits or suffers from criminal activity. In this respect, a disutility from crime can be naturally interpreted as resulting from intrinsic motivation to keep the crime level low.⁷ We show first that when the agent's degree of intrinsic motivation is not too high, then, as in the baseline model, only relatively low monitoring intensities can be implemented (which leads again to the same lower bound for the resulting crime level). In contrast, when the agent is strongly motivated to curb crime, a mirror results holds, and only relatively high monitoring intensities can be implemented. In Figure 1, these implementable sets are given by the monitoring intensities that correspond to the increasing and decreasing parts of the detection function. This gives rise to the possibility of overenforcement, which has so far not been addressed in the literature on monopolistic enforcement (see, for example, Polinsky 1980; Garoupa and Klerman 2002; Coşgel, Etkes, and Miceli 2011).⁸ However, it might occur in cases in which the harm from the offense is small or the offender's cost of not committing it is large. One example in this respect would be minor parking offenses that are fully deterred by an overly motivated agent. Moreover, intrinsic motivation to keep the number of offenses low yields an explanation for the phenomenon that there are (so many) inspectors who do monitor intensely, even if they would not suffer any material losses in case of shirking.

Fourth, depending on their degree of intrinsic motivation, agents react quite differently to incentive schemes, such as bounties for detected offenders, which are often analyzed in the literature (see, for example, Becker and Stigler 1974; Landes and Posner 1975; Polinsky 1980; Besanko and Spulber 1989; Garoupa and Klerman 2002; Coşgel, Etkes,

7. In contrast, the previous literature only considers indirect ways to induce the agent to internalize the effect of her monitoring choice on the resulting crime level. For example, Garoupa and Klerman (2002, p. 131) discuss penalties for the agent that are increasing in the number of offenses but would require the latter to be observable and verifiable. Consequently, instead of imposing penalties, Coşgel, Etkes, and Miceli (2011) argue in favor of allowing the monitoring agent to also collect income taxes (which are higher, the lower the level of crime).

8. In Landes and Posner (1975), overinvestment arises when there is a perfectly competitive market for private enforcement.

and Miceli 2011). Intuitively, agents will generate more detections when they are rewarded for doing so independent of their intrinsic motivation. However, as the number of detections is generally nonmonotonic in p (as in Figure 1), it depends on whether the inspector's optimal choice is in the increasing or in the decreasing part of the detection function. For agents with weak intrinsic motivation, the increasing part is relevant such that a higher per-detection reward indeed induces them to increase their monitoring effort. In contrast, optimal monitoring intensities of agents with strong intrinsic motivation are in the decreasing part of the detection function, so that they will reduce their effort, thereby also inducing a higher crime level. This latter result is similar to a crowding-out effect (see, for example, Deci 1971; Deci, Koestner, and Ryan 1999; Frey and Jegen 2001; Gneezy, Meier, and Rey-Biel 2011), but notice that here monetary rewards do not directly impinge on the intrinsic motivation of the agent, for example, in the sense of transforming a noneconomic relationship into an economic one (Titmuss 1970; Gneezy and Rustichini 2000a, 2000b). Rather, they simply introduce an incentive to generate more detections, which, for agents with a strong intrinsic motivation, requires a lower monitoring effort. Our analysis suggests a beneficial role of intrinsic motivation in remedying the problem of underenforcement, but it also reveals the importance of distinguishing between different types of enforcers in order to avoid severely misguided incentives.

The remainder of the paper is organized as follows: We first set up the baseline framework in Section 2, characterize implementable monitoring intensities (Section 3), and discuss some comparative-statics properties (Section 4). In Section 5, we study the case of a crime-sensitive inspector. Section 6 considers linear reward schemes (bounties), while Section 7 compares our results concerning implementability with an efficiency benchmark. Finally, Section 8 discusses some implications from our analysis and concludes. All proofs are in the Appendix.

2. MODEL

There are three types of players: a population of individuals who are potential offenders, an inspector who monitors them, and an outside governor who incentivizes the inspector. We examine each of these in turn.

2.1. Individuals

There is a unit mass of individuals who differ with respect to their gains from committing an offense, g_i , which are distributed according to a twice continuously differentiable cumulative distribution function $G: \mathbb{R} \rightarrow [0, 1]$. Following the tradition of Becker (1968), for a given probability of detection $p \in [0, 1]$ and the (exogenous) penalty $T > 0$, each individual will commit the offense if and only if its gain g_i exceeds the expected costs pT . This yields a threshold $\bar{g} := pT$ such that all individuals satisfying $g_i > \bar{g}$ ($g_i \leq \bar{g}$) will (not) commit the offense, which leads to a fraction of offenders $F(p) := 1 - G(pT)$. We assume that the distribution of the gains from the offense has full support on $[0, T]$ such that $F(p)$ is strictly decreasing.⁹

2.2. Inspector

The inspector chooses the monitoring intensity $p \in [0, 1]$ that equals the probability that each offender is detected.¹⁰ Monitoring is costly for the inspector and is captured by a strictly increasing cost function $C(p)$. Taking into account the optimal behavior of individuals as characterized above, a monitoring intensity p gives rise to a number of detected offenders $D(p) := pF(p)$. Denote by $\Delta \subseteq [0, 1]$ the image of $D(p)$ —that is, $\Delta := \{d \mid d = D(p) \text{ for some } p \in [0, 1]\}$ —and denote by p^m the smallest monitoring intensity for which the number of detections is maximal. In special cases, this occurs at the upper boundary ($p^m = 1$); otherwise, p^m is characterized by the first-order condition $D'(p^m) = 0$.

We study a context in which the inspector can be rewarded only on the basis of the number of detections $D(p)$, which is observable. Denoting the monetary reward $R[D(p)]$, the inspector's payoff is¹¹

$$u(p) = R[D(p)] - C(p). \quad (1)$$

9. Note that this assumption does not rule out the possibility that there exist individuals with $g_i < 0$ or $g_i > T$. It, however, excludes cases in which the number of offenses $F(p)$ reaches 0 for monitoring intensities $p < 1$ (as in Figure 1B). These cases could also be included in the general analysis of the model, but this would only add notational inconvenience without qualitatively affecting the results.

10. Alternatively, one could explicitly model the inspector's effort to affect the probability of detection through some (increasing) function. Under standard assumptions (for example, Inada conditions), while adding notation, this approach would not affect our results qualitatively.

11. Additive separability of rewards and costs is assumed for analytical convenience only. The assumption that the inspector's utility is not directly affected by the crime level $F(p)$ is relaxed in Section 5.

2.3. Governor

The governor remunerates the inspector by setting a payment scheme $R[D(p)]$ without being able to verify the inspector's behavior (p) or the crime level ($F(p)$).¹² In the main part of the text, it is not necessary to specify explicitly the preferences of the governor, for example, with regard to her distaste for crime. Rather, we assume that the governor aims to implement some desired monitoring intensity $\hat{p} \in [0, 1]$. For instance, \hat{p} could indeed be her privately optimal choice, or, alternatively, it could arise from a social welfare function. This latter case is discussed in Section 7.

3. IMPLEMENTABLE MONITORING POLICIES

We now analyze under which circumstances the governor can successfully induce the inspector to choose \hat{p} , that is, find payments R such that $\hat{p} \in \arg \max_p u(p)$. For any given level of detections $d \in \Delta$, define an ordered set of monitoring intensities $(P^d, <)$ such that each $p_l^d \in P^d$ satisfies $D(p_l^d) = d$. Importantly, while the number of detected offenses is equal to d for all $p_l^d \in P^d$, the underlying crime level is decreasing in l , while the inspector's effort costs are increasing in l ; that is, for all $l = 1, 2, \dots$, we have $F(p_l^d) \geq F(p_{l+1}^d)$ and $C(p_l^d) < C(p_{l+1}^d)$. Denote by P_1 the set containing all minimum monitoring intensities; that is, $P_1 = \{p \mid p = p_1^d \text{ for some } d \in \Delta\}$.

Theorem 1. A desired monitoring policy \hat{p} is implementable if and only if $\hat{p} \in P_1$. The resulting set of implementable monitoring policies P_1 satisfies $P_1 \subseteq [0, p^m]$ such that the induced number of offenses is at least $F(p^m)$.

The result is shown in Figure 2. In special cases there is a unique monitoring intensity associated with a given number of (observable) detections.¹³ Otherwise, when P^d is not a singleton, the inspector has a choice between several monitoring regimes in order to generate d detections. For example, he can choose a low level of monitoring effort p_1^d (at low cost), which leads to a relatively high number of offenders, out of which d are

12. See Arlen (1994), Garoupa and Klerman (2002), and Coşgel, Etkes, and Miceli (2011) for similar assumptions concerning the governor's role as an outsider in the sense of lacking these crucial pieces of information.

13. In the example in Figure 2, this is true for the global maximum $D(p^m)$ and when the number of detections is very small.

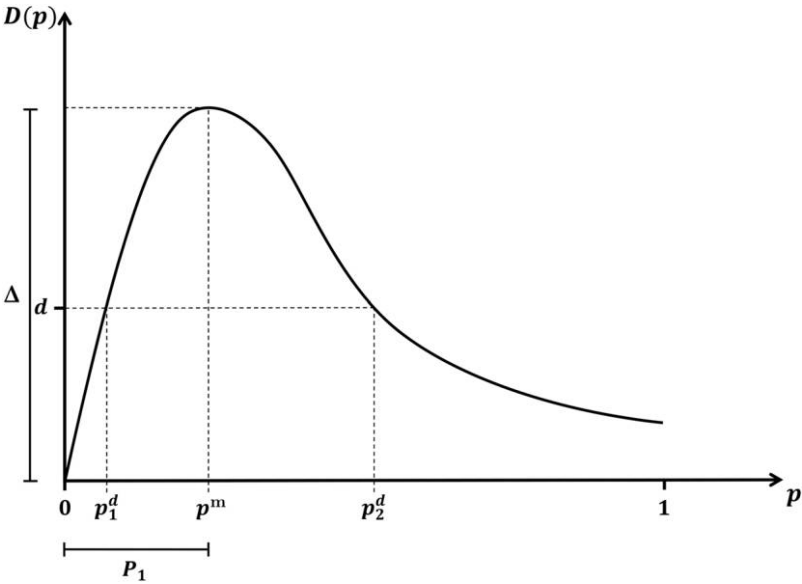


Figure 2. Number of detections as a function of monitoring intensity p

detected (scapegoats). Alternatively, the inspector can choose a higher level of effort $p_2^d > p_1^d$ (at higher cost), which leads to fewer offenses but again to d detections (black sheep). Since the governor can observe only the number of detections but not the chosen monitoring intensity, these two choices of effort are observationally identical from the governor's point of view. As the inspector's payment is the same for all $p_i^d \in P^d$, he prefers to deliver any given number of detections $d \in \Delta$ at the lowest cost, and so his optimal choice is p_1^d . Consequently, only monitoring policies $p \in P_1$ can be implemented so that $\hat{p} \in P_1$ is a necessary condition for its implementation. As for sufficiency, all monitoring levels $\hat{p} \in P_1$ can be implemented by sufficiently rewarding the corresponding detection level $D(\hat{p})$, compared with all other detection levels $d \neq D(\hat{p})$.

In a next step, we analyze in more detail the set of implementable monitoring policies P_1 . First, since theorem 1 renders all $p > p^m$ nonimplementable, the crime level is bounded from below by $F(p^m)$. Therefore, P_1 's upper-bound p^m becomes crucial. The value of p^m is determined by the shape of the detection function $D(p)$, and its hump-shaped representation in Figure 2 is quite characteristic. To see this, note that it always holds that $D(0) = 0$ and $D'(0) = F(0) > 0$ since without monitoring there

are no detections (but a high crime level). Moreover, full monitoring typically also leads to few detections ($D(1) = 1 - G(T)$), because, given that every offense is detected, there are not many offenders (that is, only those who are undeterrable as $g_i > T$). Thus, the maximal number of detections $D(p^m)$ is usually attained between these two extremes.

Second, even some $p < p^m$ might be nonimplementable, in which case P_1 would be only a strict subset of the interval $[0, p^m]$. This case occurs when $D(p)$ is not monotonically increasing over this interval, such that there would exist several monitoring intensities leading to the same number of detections. By theorem 1, only the minimum of these can be implemented. Otherwise, when $D(p)$ is monotonically increasing over $[0, p^m]$, we have $P_1 = [0, p^m]$ such that a monitoring intensity p is implementable if and only if $p \leq p^m$. The two properties discussed above— p^m interior and $D(p)$ monotonically increasing over $[0, p^m]$ —can be traced back to the distribution of gains from crime such that we obtain corollary 1.

Corollary 1. For the set of implementable monitoring policies P_1 , the following hold:

i) Let the number of undeterrable individuals be sufficiently small such that it satisfies the condition $1 - G(T) < TG'(T) (> 0)$. Then p^m is interior with the consequence that not all monitoring policies are implementable; that is, $P_1 \subsetneq [0, 1]$.

ii) Let G be not too concave such that it satisfies the condition $G''(g) > -G'(g)/g (< 0)$ for $g \in [0, T]$. Then the number of detections $D(p)$ is monotonically increasing over $[0, p^m]$ with the consequence that a desired monitoring intensity \hat{p} is implementable if and only if $\hat{p} \leq p^m$; that is, $P_1 = [0, p^m]$.

Corollary 1 provides two, arguably mild, conditions that are jointly sufficient for the detection function to be hump shaped on its domain $[0, 1]$. Both statements of corollary 1 are derived from the detection function $D(p) = p[1 - G(pT)]$. Writing the first-order condition $D'(p) = 0$ as

$$1 - G(pT) = pTG'(pT) \quad (2)$$

reveals the two underlying marginal effects: The term on the left-hand side captures the higher number of detections as the monitoring intensity increases (for a given crime level). The term on the right-hand side measures the marginal deterrence effect (for a given probability of detection). Condition i of corollary 1 ensures that there exists an interior monitoring intensity p that satisfies equation (2), that is, that balances the two

marginal effects. This is achieved by simply requiring that for $p = 1$ the marginal detection effect (left-hand side) is smaller than the marginal deterrence effect (right-hand side), which is never true for $p = 0$. Since by theorem 1 only $p \leq p^m$ are implementable, under condition i of corollary 1 there exist monitoring intensities that are not.¹⁴ Condition ii of corollary 1 ensures that the right-hand side of equation (2) is increasing in p ; that is, the marginal deterrence increases as monitoring becomes more and more intense.¹⁵ Since the left-hand side of equation (2) is always decreasing in p , when condition ii of corollary 1 holds, there is at most one monitoring intensity p that solves equation (2). This implies that the slope of the detection function changes its sign at most once (that is, turns from increasing to decreasing since $D'(0) = F(0) > 0$). Therefore, under condition ii of corollary 1, $D(p)$ must be monotonically increasing for $p < p^m$, and, hence, all of these monitoring intensities are implementable.

Remark. Theorem 1 can also be expressed in terms of the elasticity of crime $\varepsilon(p) := -F'(p)(p/F(p))$. Under conditions i and ii of corollary 1, it is readily derived that $\varepsilon(p) \geq 1$ if and only if $p \geq p^m$. Thus, as a rule, inspectors cannot be induced to choose a monitoring regime in the elastic range of the crime function.

4. COMPARATIVE-STATICS ANALYSIS

The implications from theorem 1 depend strongly on the model's fundamentals, in particular the underlying distribution of the gains from misbehavior G and the fine T . In this section we use comparative-statics analysis to assess the impact of these two factors.

4.1. Impact of the Distribution of Gains

Different distributions of gains from crime G give rise to varying levels of misbehavior, detections, and implementable monitoring intensities. We compare distributions that differ in the sense of first-order stochastic dominance (FOSD). To this end, consider two distributions G and \tilde{G} , where G is first-order stochastically dominated by \tilde{G} ; that is,

14. Polinsky (1980) provides further conditions that are sufficient for p^m interior such that monitoring policies close to 1 cannot be implemented.

15. In fact, corollary 1.ii even implies that $D(p)$ is concave. Together with corollary 1.i, which guarantees that p^m is interior, this ensures that $D(p)$ is hump shaped on $[0, 1]$, similar to its illustration in Figure 2.

$G(g) \geq \tilde{G}(g)$ for all g , with the interpretation that \tilde{G} has more probability mass on high gains from crime. Denote by \tilde{p}^m and \tilde{P}_1 the respective maximizer of the number of detections and the set of implementable monitoring policies resulting under \tilde{G} .

Proposition 1. Let G be first-order stochastically dominated by \tilde{G} .

i) Then, for any given monitoring intensity p , the number of detections and the level of misbehavior are larger under \tilde{G} than under G .

ii) Let G and \tilde{G} satisfy conditions i and ii of corollary 1. If the slopes of \tilde{G} and G at $p = p^m$ are not too distinct, then the set of implementable monitoring intensities is larger under \tilde{G} than under G . Formally, if

$$\tilde{G}'(p^m T) - G'(p^m T) < \frac{G(p^m T) - \tilde{G}(p^m T)}{p^m T} (\geq 0),$$

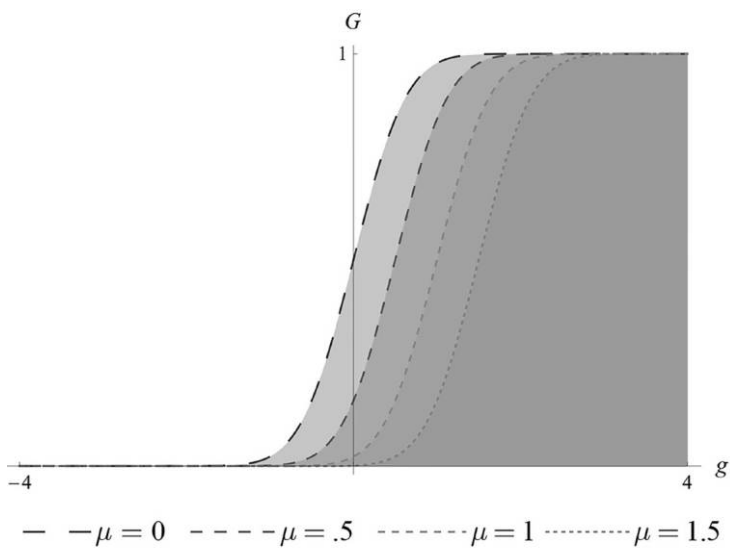
then $\tilde{p}^m > p^m$, and hence $\tilde{P}_1 \supsetneq P_1$.

The two parts of proposition 1 suggest that there is a trade-off in the sense that facing a population with a low tendency toward misbehavior (as exemplified by distribution G) is on the one hand beneficial, as the level of misbehavior is relatively low for any given monitoring intensity p (proposition 1.i). But on the other hand, only a small set of (low) monitoring intensities is implementable, which in turn might still lead to relatively high levels of misbehavior (proposition 1.ii).

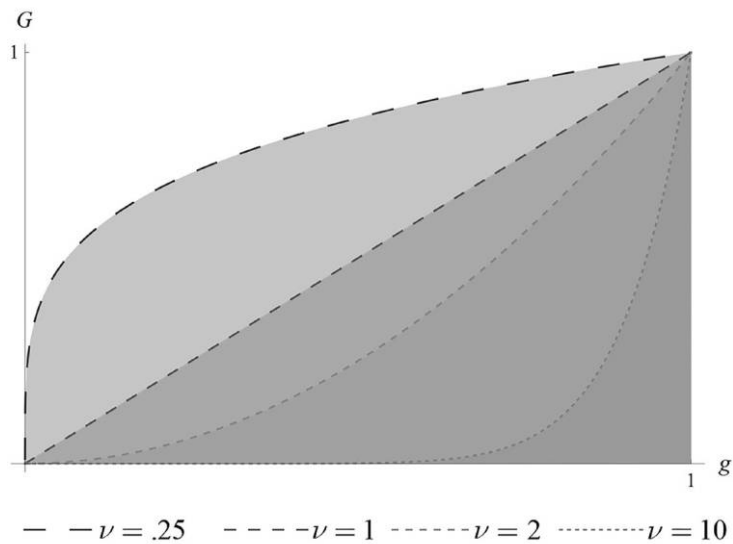
How these two effects unfold is demonstrated in Figures 3–5 for two classes of distributions (and for $T = 1$): normal distributions $N(\mu, .5)$ and power distributions with cumulative distribution functions (CDFs) $G(g) = g^\nu$ (defined for $g \in [0, 1]$). Shown are the CDFs (Figure 3), the detection function $D(p)$ (Figure 4), and the crime function $F(p)$ (Figure 5) for the different parameter values, where those distributions with higher parameter values of μ and ν first-order stochastically dominate those with lower values.

Figures 4 and 5 illustrate the trade-off as emerging from proposition 1: when there is large probability mass on low realizations of g (low values of μ and ν), the overall levels of misbehavior and detections are low. But in addition, the value p^m where the (hump-shaped) detection function $D(p)$ reaches its peak is small, so that, from corollary 1, the set of implementable monitoring intensities $P_1 = [0, p^m]$ is relatively small.

Shifting probability mass to higher realizations of g (that is, increasing μ and ν) then leads to upward shifts of both $F(p)$ and $D(p)$ since deterrence is now weaker for any level of p . As a result, p^m and hence also the

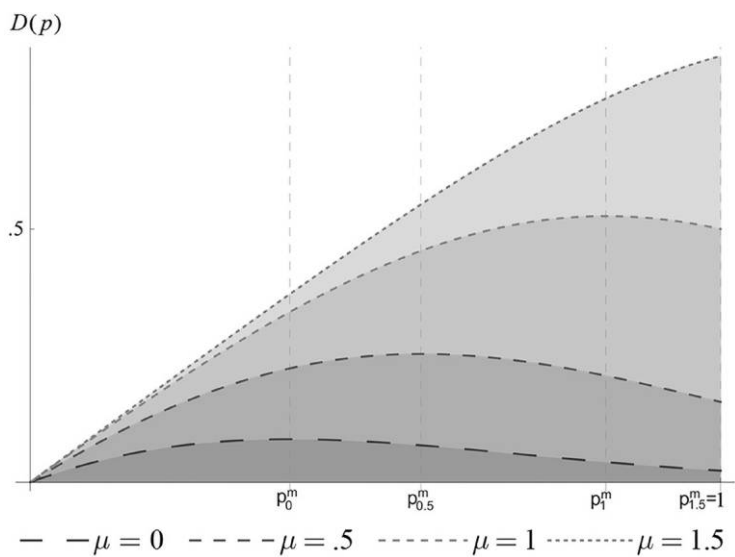


(A) Normal distribution

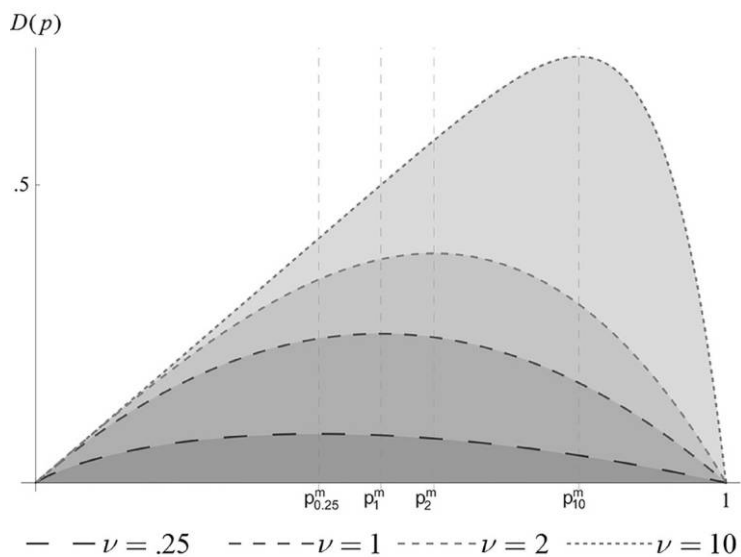


(B) Power distribution

Figure 3. Different classes of distributions: cumulative distribution functions

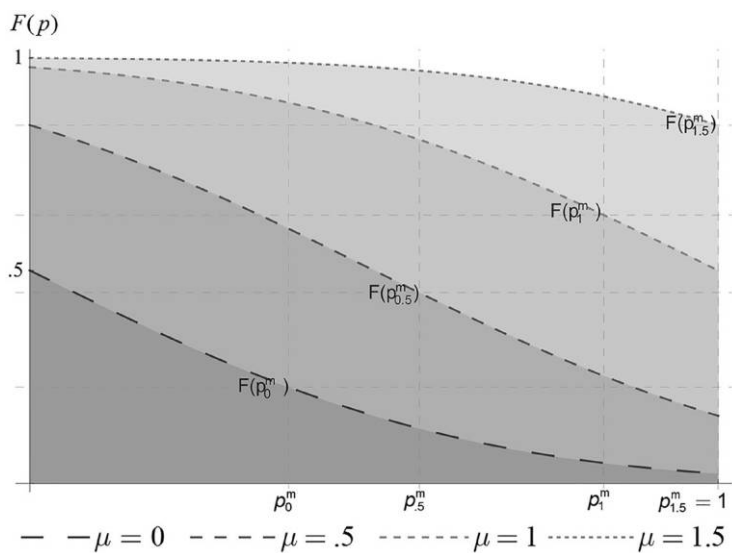


(A) Normal distribution

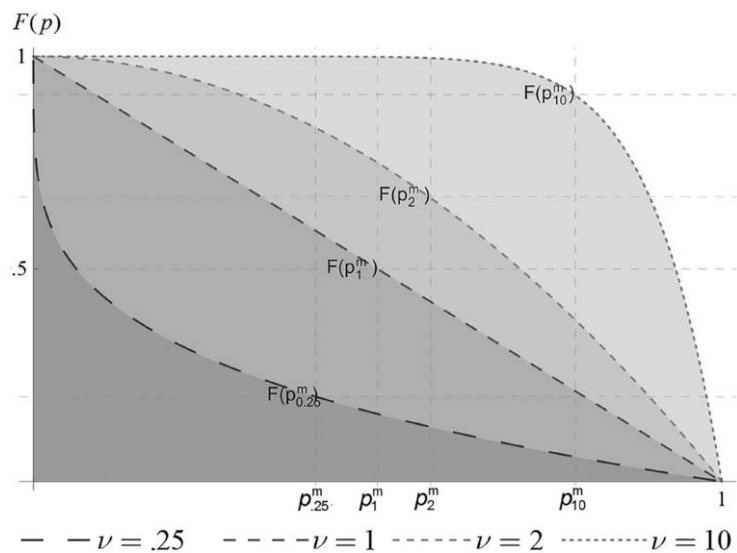


(B) Power distribution

Figure 4. Different classes of distributions: detections



(A) Normal distribution



(B) Power distribution

Figure 5. Different classes of distributions: crime levels

upper bound of the set P_1 increase.¹⁶ When there are sufficiently many individuals with large gains from the offense, $D(p)$ may even become monotonically increasing on $[0, 1]$ as in Figure 4A, so that $P_1 = [0, 1]$, thereby coinciding with the choice set of the inspector. In this case, condition i of corollary 1 is violated; that is, there are many undeterrable individuals, and there is no loss in terms of nonimplementable monitoring intensities, so that theorem 1 has no bite.

Interestingly, with respect to the resulting minimum crime level $F(p^m)$, the benefit from a larger choice set P_1 due to a larger p^m can outweigh the cost in the form of an upward shift of the crime function $F(p)$. For example, for the case of the normal distribution with the monotonically increasing detection function ($\mu = 1.5$) just discussed, there are many criminals who are not deterred even when monitored with the maximum feasible intensity $p = p^m = 1$ (see Figures 4A and 5A). Therefore, the minimum crime level is $F(1) = .84$. In contrast, for the power distributions considered, full deterrence is in principle possible as $F(1) \equiv 0$ for all $\nu > 0$ (see Figure 5B). However, only monitoring intensities $p \in [0, p^m]$ are implementable, and, as can be seen from Figure 4B, $p^m < 1$ holds. Therefore, the minimum crime level $F(p^m)$ might well exceed that resulting under the monotonic case: for example, for $\nu = 10$, we get $p^m = .78 < 1$, which leads to $F(p^m) = .91 > .84$.

4.2. Varying the Fine

To investigate the impact of changes in the exogenous fine T , it is useful to now express explicitly, where appropriate, the dependency on T ; that is, we now write $F(p, T)$, $D(p, T)$, and $p^m(T)$. Analogous to the comparative statics concerning the distribution G (embodied in proposition 1), changes in the fine T have two countervailing effects on the set of implementable monitoring policies.

Proposition 2. Consider two fines T and \tilde{T} with $T < \tilde{T}$.

- i) For any given monitoring intensity $p > 0$, the number of detections and the level of misbehavior are smaller under \tilde{T} than under T ; that is, $F(p, \tilde{T}) < F(p, T)$ and $D(p, \tilde{T}) < D(p, T)$.
- ii) Let G satisfy conditions i and ii of corollary 1. Then the set of im-

16. For example, consider the case $\nu = 1$ in the power distribution, which corresponds to the uniform distribution of gains over $[0, 1]$. Since $p^m = \frac{1}{2}$, the crime level is at least $F(p^m) = \frac{1}{2}$. Thus, when the governor wants to deter more than half of the population (for instance, because the harm from the offense is large), then the required monitoring policy $\hat{p} > \frac{1}{2}$ cannot be implemented.

plementable monitoring intensities is smaller under \tilde{T} than under T . Formally, $\tilde{P}_1 \supsetneq P_1$ because $p^m(\tilde{T}) < p^m(T)$.

Proposition 2.i shows that an increase in T leads to a downward shift of $F(p, T)$ and $D(p, T)$, thereby unambiguously lowering both the number of offenses and the number of detections. Intuitively, for every given p , a higher fine increases the expected penalty from committing the offense and hence leads to fewer offenses and, as a result, fewer detections.

However, analogous to the comparative statics concerning the distribution G , proposition 1.ii points to a countervailing effect in the sense that it leads to a smaller set of implementable monitoring intensities $P_1 = [0, p^m(T)]$. This result is in contrast to the celebrated finding of Becker (1968), according to which any expected fine pT should be implemented with T as large as possible, as increasing T is costless (in contrast to increasing p being costly). For the contexts considered here, this suggests that there is a reduced benefit associated with increasing T in the form of a shrinking set of implementable monitoring intensities. This trade-off is illustrated in Figure 6 for the case in which the gains from crime are distributed according to a standard normal distribution $N(0, 1)$ and for different levels of T . As can be seen in Figure 6, low values of T give rise to a large set of implementable monitoring intensities (that is, $p^m(T)$ is large), but the resulting crime level (and the number of detections) is high. A higher value of T reduces the crime level, but it also reduces the set of monitoring intensities that can be implemented by the governor. Which effect dominates depends on whether $p^m(T)T$ is increasing or decreasing in T , which in turn depends again on the underlying distribution G . Interestingly, in the example of the normal distribution depicted in Figure 6, the two effects just balance each other such that the minimum crime level $F(p^m(T), T)$ remains constant. For the case of power distributions, this property can even be shown analytically: that is, as long as $p^m(T)$ is interior, $F(p^m(T), T)$ is independent of T .¹⁷ This implies that, in contrast to standard arguments, the minimum crime level cannot be reduced by an increase in the fine T .

17. Indeed, consider $F(p, T) = 1 - (pT)^\nu$ for $pT \leq 1$. Using the first-order condition, we obtain $p^m(T) = 1/[(\nu + 1)^{1/\nu}T]$, which is interior given that $T > (1 + \nu)^{-1/\nu}$. Finally, $F(p^m(T), T) = 1 - [p^m(T)]^\nu \times T^\nu = \nu/(1 + \nu)$.

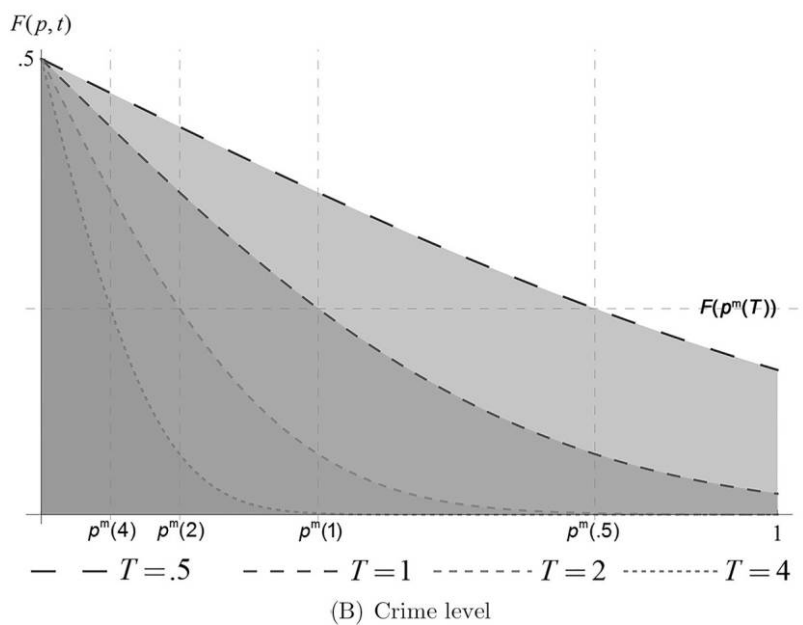
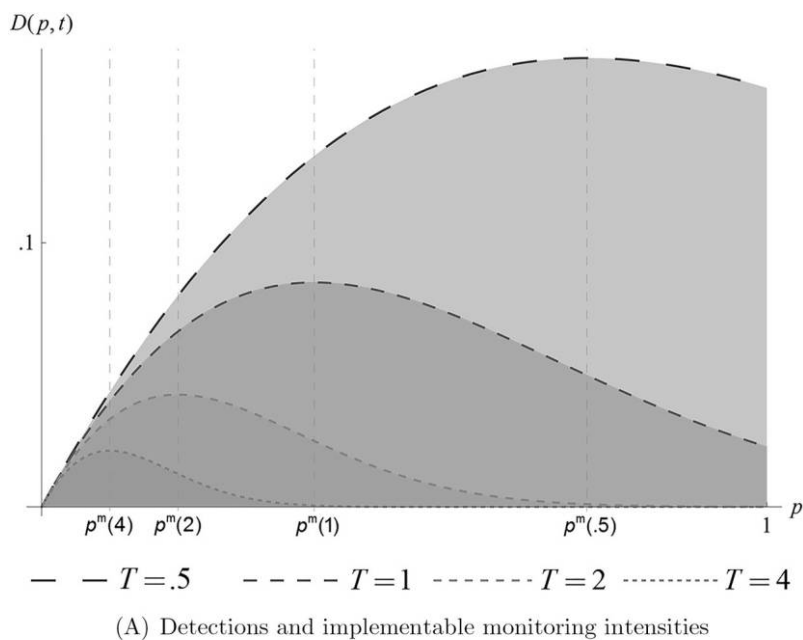


Figure 6. Effect of changes in the fine T

5. CRIME-SENSITIVE INSPECTORS

One major implication of theorem 1 is that inspectors cannot be induced to choose monitoring intensities beyond the ceiling p^m , which puts a lower bound $F(p^m)$ on the resulting crime level. We now investigate to what extent this result relies on the assumption maintained so far that inspectors care only about their remuneration and costs of effort and not about the level of crime itself (see equation [1]). In this section, we relax that assumption and allow for crime-sensitive inspectors, characterized by the more general utility function

$$u(p, \beta) = R[D(p)] - \beta F(p) - C(p), \quad (3)$$

where β measures how the inspector's utility is affected by the presence of crime. Thereby, disutility ($\beta > 0$) can be caused by intrinsic motivation to keep the crime level low, while benefits of crime ($\beta < 0$) can arise, for example, from accepting bribes.¹⁸ The inspector's preferences from the basic model are nested as the special case $\beta = 0$ in this utility function.

As will be shown, inspectors with a sufficiently strong intrinsic motivation can be induced to implement monitoring intensities $p > p^m$, which are not implementable in the basic setup. However, it turns out that for any type of inspector, there is still a potentially large set of monitoring intensities that are not implementable.

To illustrate this point, we set T equal to 1 and consider quadratic costs of effort $C(p) = cp^2$ for some cost parameter $c > 0$. Moreover, for the sake of analytical tractability, we focus on the case in which the gains from the offense are uniformly distributed on $[0, 1]$, which corresponds to the special case $\nu = 1$ for the power distributions considered in Figure 3, leading to $F(p) = 1 - p$. The resulting detection function $D(p) = p(1 - p)$ is then hump shaped and symmetric around $p^m = \frac{1}{2}$. Also, the conditions of corollary 1 are satisfied, so that the set of implementable monitoring intensities is $P_1 = [0, \frac{1}{2}]$ for an inspector who is not crime sensitive ($\beta = 0$). The following result characterizes this set for crime-sensitive inspectors ($\beta \neq 0$).

18. The literature discusses other reasons why inspectors might worry about the prevailing level of crime, for example, payments that are inversely related to the crime level (Garoupa and Klerman 2002) or a lower tax revenue with part accruing to the enforcer (Coşgel, Etkeş, and Miceli 2011). In a different framework, Besanko and Spulber (1989) assume that an inspector can allocate a given budget between enforcement and perquisites, so that the marginal rate of substitution can also be interpreted as a measure of the inspector's concern about crime.

Proposition 3. Let the distribution of gains be uniform on $[0, 1]$ and $T = 1$ such that $F(p) = 1 - p$ and $p^m = \frac{1}{2}$. Moreover, let the inspector's utility function be given by equation (3) with $C(p) = cp^2$. When the inspector's disutility from crime is sufficiently high (that is, for $\beta > c$), then the set of implementable monitoring policies is $[\frac{1}{2}, 1]$. Otherwise (that is, for $\beta < c$), it is given by $P_1 = [0, \frac{1}{2}]$.

The underlying intuition for the proposition is simple: any reward offered for the desired number of detections $D(\hat{p})$ can also be gained with mimicking \hat{p} by choosing $\tilde{p} = 1 - \hat{p}$ instead since $D(\tilde{p}) = D(\hat{p})$. It then depends on the inspector's sensitivity to crime β , in relation to the costs of monitoring, whether the higher or the lower monitoring intensity is preferred. In the knife-edge case of $\beta = c$, the inspector is indifferent between each pair of monitoring intensities that lead to the same number of detections.

Proposition 3 suggests that inspectors can be distinguished with respect to their β type as follows: For bad types ($\beta < c$), the intrinsic motivation for curbing crime is low, and we are back to the setting underlying theorem 1, which renders monitoring intensities $p > p^m$ nonimplementable. The resulting crime level tends to be high (that is, larger than $F(p^m)$), and detected offenders should hence be classified as scapegoats.

In contrast, for good types ($\beta > c$), the intrinsic motivation is sufficiently high such that they can be induced to choose high monitoring intensities. In this case, the number of offenses is low, and the detected offenders should rather be classified as black sheep. Notice, however, that in this case, a mirror result of theorem 1 applies in the sense that only monitoring intensities $p \geq p^m$ can be implemented, and lower ones cannot.

6. LINEAR REWARD SCHEMES

The inspector's degree of intrinsic motivation also crucially determines his behavioral response to changes in the reward scheme $R[D(p)]$. We illustrate this effect by considering simple linear payment schemes in which for each detection the inspector receives a predefined reward (or bounty) r ; that is, $R[D(p)] = rD(p)$. This specification of payments is employed in the literature on delegated (private) enforcement (see, for example, Becker and Stigler 1974; Landes and Posner 1975; Polinsky 1980; Besanko and

Spulber 1989). The inspector's optimal monitoring policy $p^*(r, \beta)$ then satisfies

$$p^*(r, \beta) \in \arg \max_{p \in [0, 1]} rD(p) - \beta F(p) - C(p).$$

Keeping the parameterizations of Section 5, this leads to the following result:

Proposition 4. Let the distribution of gains be uniform on $[0, 1]$ and $T = 1$ such that $F(p) = 1 - p$, and let the inspector's utility function be given by equation (3) with $C(p) = cp^2$. If $R[D(p)] = rD(p)$ with r not too small (that is, $r \geq \max\{-\beta, \beta - 2c\}$), then

- i) the inspector's optimal monitoring policy is given by $p^*(r, \beta) = (r + \beta)/2(r + c)$, which is increasing in β , and
- ii) an increase in the reward r leads bad (good) types to optimally choose a higher (lower) monitoring intensity; that is, $\partial p^*(r, \beta)/\partial r > 0 \Leftrightarrow \beta < c$ and $\partial p^*(r, \beta)/\partial r < 0 \Leftrightarrow \beta > c$.

As for proposition 4.i, more intrinsic motivation to curb crime β leads to a higher level of monitoring effort.¹⁹ Thereby, $p^*(r, \beta) > p^m = \frac{1}{2}$ if and only if $\beta > c$, so that indeed each type optimally chooses a monitoring intensity from the set of implementable ones as characterized in proposition 3.

Proposition 4.ii reveals that the responses of the two types of inspectors to changes in the bounty r are quite different: For bad types ($\beta < c$), starting at the lower bound $r = -\beta$ we have $p^*(r, \beta) = 0$, and the monitoring intensity increases as r increases, approaching $p^*(r, \beta) = \frac{1}{2}$ from the left in the limiting case $r \rightarrow \infty$.²⁰ In this case, higher monetary rewards have the standard effect of increasing the monitoring intensity, generating more detections, and thereby lowering the number of offenses.²¹

19. In the framework of Coşgel, Etkes, and Miceli (2011), the inspector chooses a higher level of effort when he benefits from a lower level of crime in the form of higher taxation income.

20. This is shown formally at the end of the proof of proposition 4; see also Garoupa and Klerman (2002).

21. In the context of corporate criminal liability considered in Arlen (1994), a firm is required to monitor its employees, and because of vicarious liability, it is liable for crimes committed by them. Liability is triggered (only) when a crime is detected by the firm so that, similar to our framework, the introduction of vicarious liability (which can be interpreted as a negative bounty for each detection) may in fact decrease the firm's optimal monitoring intensity.

In contrast, for good types ($\beta > c$), $p^*(r, \beta)$ is decreasing in the reward r , starting at the lower bound $r = \beta - 2c$, where $p^*(r, \beta) = 1$ and approaches $p^*(r, \beta) = \frac{1}{2}$ from the right as r increases. Note that these types also respond by generating more detections when the marginal benefit r of doing so increases. However, they achieve more detections by reducing their monitoring intensity, which leads to more offenses. This result is in line with the literature on motivational crowding out in the sense that the inspector's monitoring effort is lower the stronger the monetary incentives (see, for example, Deci 1971; Deci, Koestner, and Ryan 1999; Frey and Jegen 2001; Gneezy, Meier, and Rey-Biel 2011). In our framework, however, this effect is not due to the fact that such incentives directly reduce the inspector's degree of intrinsic motivation, for example, in the sense of turning a noneconomic relationship into an economic one (see, for example, Titmuss 1970; Frey and Oberholzer-Gee 1997; Gneezy and Rustichini 2000a, 2000b). Rather, for inspectors with a high degree of intrinsic motivation, generating more detections requires a lower monitoring effort, thereby also leading to more offenses.

The dependence of the inspector's response to monetary rewards on his degree of intrinsic motivation to keep the crime level low has further consequences: for example, in order to implement some desired monitoring intensity $\hat{p} > \frac{1}{2}$, the corresponding bounty is $r(\hat{p}, \beta) = (\beta - 2\hat{p}c)/(2\hat{p} - 1)$, which is always positive when the inspector's intrinsic motivation is sufficiently strong (that is, for $\beta > 2c$), but $r(\hat{p}, \beta) < 0$ (that is, a payment from the inspector to the governor for each detection) is also possible. This occurs for good types with intermediate degrees of intrinsic motivation ($c < \beta < 2c$) and when $\hat{p} > p^*(0, \beta)$, that is, when \hat{p} exceeds the monitoring intensity chosen by the inspector under intrinsic motivation alone ($r = 0$). In that case, the inspector needs to be punished by a negative bounty in order to increase his effort.

One can also relax the assumption that the fine is fixed at $T = 1$ and explore how changes in T affect the per-detection reward necessary to induce a given implementable monitoring intensity as optimally chosen by the inspector. For r such that the inspector's choice is interior, we have $r(\hat{p}, \beta, T) = (2c\hat{p} - \beta T)/(1 - 2T\hat{p})$. For the case of purely monetary incentives ($\beta = 0$), this is increasing in T so that higher fines induce higher per-detection rewards. With crime-sensitive inspectors ($\beta \neq 0$), however, the sign of the effect is ambiguous. Taking the derivative with respect to T reveals that it delicately depends on the parametrization of the model

whether rewarding the inspector for detections becomes more or less costly for the governor as the fine increases.²²

All in all, our findings suggest that payment schemes based on the number of detections such as bounties are delicate instruments that are sensitive to the inspector's concern for keeping the crime level low. In particular, ignoring an inspector's intrinsic motivation might lead to severely misguided incentives.

7. SOCIAL WELFARE

So far, the analysis has focused on the implementability of a desired monitoring intensity (\hat{p}) that is assumed to be exogenously given. In this section we relax this assumption and derive \hat{p} endogenously as the solution to a maximization problem of the governor who can be considered a social planner. This provides us with an efficiency benchmark against which we then compare our results for implementability. In this respect, the underlying social welfare function contains the following elements. First, let the society's harm from crime be $hF(p)$, where $h > 0$ is a constant harm from each offense. Furthermore, from the inspector's payoff function (see equation [3]), we include the inspector's enforcement costs $C(p)$ and his disutility from crime $\beta F(p)$, while the remuneration $R[D(p)]$ is a transfer payment between the governor and the inspector, which cancels. Moreover, the offenders' gains from the offense are typically included (see, for example, Polinsky and Shavell 2007), but this is not an uncontroversial issue (see, for example, Stigler 1970). For this reason, we use a general specification that nests both possibilities as special cases, so that social welfare is given by

$$SW(p) := -(h + \beta) \times F(p) - C(p) + \pi \int_{pT}^{\infty} g \times G'(q) dq, \quad (4)$$

where the parameter $\pi \in [0, 1]$ reflects the weight that the social planner puts on the offenders' gains from the offense.²³

22. A special case of a linear reward occurs when the inspector keeps the collected fines; that is, $r = T$. If interior, the optimal monitoring intensity is then given by $p(\beta, T) = (1 + \beta)T/(2T^2 + 2c)$, which is hump shaped in T where the maximum is attained at $T = \sqrt{c}$. Hence, for fines beyond this threshold, a higher fine (that is, a higher reward per detection) would induce the inspector to choose a lower monitoring intensity.

23. The special case $\pi = 0$ is consistent with an alternative interpretation. If the inspector has an outside option of 0 and the governor chooses the reward to satisfy the inspector's participation constraint, then $R[D(p)] = \beta F(p) - C(p)$, such that the governor's payoff coincides with social welfare defined in equation (4), when setting π equal to 0.

Throughout, we continue with the uniform-quadratic specification; that is, the gains from the offense are uniformly distributed on $[0, 1]$, $T = 1$ (so that $F(p) = 1 - p$ and $p^m = \frac{1}{2}$) and $C(p) = cp^2$. The desired monitoring intensity \hat{p} can thus be defined as maximizing the following social welfare function:

$$\hat{p} = \arg \max_{p \in [0, 1]} SW(p) = (1 - p) \left[\frac{\pi}{2} (1 + p) - h - \beta \right] - cp^2. \quad (5)$$

It is easily seen that social welfare is always strictly concave in p and, given that $h + \beta > 0$, also strictly increasing at $p = 0$ such that the welfare optimum either is attained at the boundary ($\hat{p} = 1$), which here leads to full deterrence, or is interior.²⁴ In the latter case, it solves the first-order condition $SW'(\hat{p}) = 0$ and is given by

$$\hat{p} = \frac{h + \beta}{\pi + 2c}. \quad (6)$$

Hence, the efficient (interior) monitoring intensity increases as offenses become more detrimental for the society (h), when the inspector becomes more crime sensitive (β), as the gains from the offense receive less weight in the social welfare function (π), and when the cost of monitoring decreases (c).

In a next step, we compare the efficiency benchmark with our previous findings concerning implementability (in particular, theorem 1 and proposition 3). This leads to the following result:

Proposition 5. Let social welfare be given by equation (5) so that the resulting socially optimal monitoring intensity is \hat{p} .

i) When the inspector's crime sensitivity is low ($\beta < c$; bad type), \hat{p} is implementable if and only if the harm from the offenses is sufficiently small; that is, $h \leq c - \beta + \pi/2$. Otherwise, underenforcement arises.

ii) When the inspector's crime sensitivity is high ($\beta > c$; good type), \hat{p} is implementable if and only if the harm from the offenses is sufficiently large; that is, $h \geq c - \beta + \pi/2$. Otherwise, overenforcement arises.

The proposition can be illustrated using the taxonomy presented in Table 1. Thereby, p^* denotes the (privately) optimal monitoring intensity

24. The socially optimal monitoring intensity is interior for sufficiently large costs of monitoring. More precisely, given that $h + \beta > 0$, \hat{p} is interior if and only if $c > (h + \beta - \pi)/2$.

Table 1. Privately Optimal versus Efficient Monitoring Intensities

	Low Harm from Crime: $b < c - \beta + \frac{\pi}{2}$	High Harm from Crime: $b > c - \beta + \frac{\pi}{2}$
Low sensitivity to crime: $\beta < c$	$p^*, \hat{p} \leq p^m$ Implementable	$p^* \leq p^m < \hat{p}$ Underenforcement
High sensitivity to crime: $\beta > c$	$\hat{p} < p^m \leq p^*$ Overenforcement	$p^*, \hat{p} \geq p^m$ Implementable

chosen by the inspector for a given reward scheme $R[D(p)]$; that is, p^* maximizes the inspector’s payoff as given in equation (3).²⁵

As for proposition 5.i, bad types exhibit a low intrinsic motivation to curb crime such that only p^* from the set $[0, p^m]$ can be implemented (see proposition 3). If the level of social harm from crime is sufficiently low, the welfare-maximizing monitoring intensity \hat{p} also belongs to this set and hence becomes implementable. For example, it can be shown that by setting a bounty $\hat{r} = [ch - (\beta\pi/2)]/[(\pi/2) - b - \beta + c]$, the governor indeed induces the inspector to choose $p^* = \hat{p}$.²⁶ For greater harm from crime, $p^* \leq p^m < \hat{p}$ holds, which implies that underenforcement is inevitable.

Analogously, in proposition 5.ii the inspector’s intrinsic motivation level is high, so that only high monitoring intensities $p^* \in [p^m, 1]$ can be implemented (see proposition 3). Hence, the welfare-maximizing monitoring intensity also belongs to this set only when the harm from crime is sufficiently high such that $\hat{p} > p^m$ holds. For low levels of harm from crime, overenforcement arises since $\hat{p} < p^m \leq p^*$. Note that the conditions for overenforcement require that $b < \pi/2$ (since we must have $\beta + b - (\pi/2) < c < \beta$), so that overenforcement can occur only when the offenders’ gains from the offense also enter the social welfare function ($\pi > 0$).

25. Note that there typically exist several reward schemes that lead to the same p . For example, in the proofs of theorem 1 and proposition 3, we consider a simple discontinuous reward scheme. But it can be shown that, except for p^m (where the detection function $D(p)$ reaches its hump), all implementable monitoring policies can also be reached with a linear reward scheme, as discussed in Section 6.

26. Importantly, in our model there are no restrictions on the feasible transfers between the governor and the inspector (for example, due to limited liability). This implies that in our setup there is no issue with additional distortions because of a trade-off between rent and efficiency, which arises in many agency models (see, for example, Laffont and Martimort 2002) and would affect the possible implementability of \hat{p} even on the diagonal of Table 1.

In terms of this taxonomy, one could argue that the previous literature (see, for example, Polinsky 1980; Garoupa and Klerman 2002; Coşgel, Etkes, and Miceli 2011) confines attention to the cases with low sensitivity to crime in Table 1, while we also analyze the cases with high sensitivity to crime.²⁷ We consider both of these novel cases to be empirically relevant. Intrinsic motivation might be one explanation why in many situations where the desired level of monitoring intensity is high there is no issue of implementability, although an inspector's shirking would not reduce his payments. And intrinsic motivation can cause overenforcement when the governor's desired monitoring level is low, for example, in the case of minor offenses that are fully deterred by an overly motivated inspector.

In general, our analysis points to some degree of congruity of the inspector's crime sensitivity (β) and the social harm from crime (b), which is required in order to make the efficient monitoring intensity implementable. This is satisfied on the diagonal of Table 1. Hence, when the level of harm is high, inspectors with strong intrinsic motivation are socially desirable in order to overcome the underenforcement issue. In contrast, when the level of harm from crime is low, inspectors with low levels of intrinsic motivation seem better suited since they do not overenforce.²⁸

It is also interesting to investigate how a change in the fine T would affect these results. While a full analysis is beyond the scope of the paper, it is clear that the following effects need to be taken into account. First, as shown in proposition 2, the monitoring intensity for which the number of detections is maximized becomes smaller as the fine increases (that is, $p^m(T)$ is decreasing in T), which shrinks the set of implementable monitoring intensities for inspectors with low levels of intrinsic motivation (bad

27. Indeed, propositions 1 and 5 in Polinsky (1980) refer to high crime and underenforcement; Garoupa and Klerman (2002) show that there is a linear reward scheme (bounty) that implements a desired monitoring policy under low levels of harm but not high levels of harm; and Coşgel, Etkes, and Miceli (2011) introduce an inspector's crime sensitivity by the right to collect taxes, but they do not explore the possibility that the disutility from crime is strong enough that the lower cells become relevant.

28. The optimal alignment of preferences in principal-agent settings with delegation is also discussed in contexts other than law enforcement. For instance, Bubb and Warren (2014) study a principal's optimal delegation decision when agents differ with respect to how strongly they care about the social costs and benefits from a regulatory policy such as environmental protection. They show that it is typically not optimal for the principal to select an unbiased agent (that is, one with the same preferences as the principal with respect to the policy). The reason is that biased agents might have a stronger incentive to engage in information acquisition about new regulatory opportunities.

types). Second, a higher fine changes the threshold for the classification as either bad or good types (now the condition is $\beta \leq c/T^2$ instead of $\beta \leq c$). Third, the fine also affects the welfare-maximizing (interior) monitoring intensity $\hat{p} = [T(b + \beta)]/(\pi T^2 + 2c)$, which is increasing in T as long as the weight of the gains from the offenses in the social welfare function (π) is not too large. Taken together, whether a higher fine increases the scope of implementability of \hat{p} or instead fosters underenforcement or overenforcement is generally ambiguous and depends on the underlying distributional assumptions as well as on the model's other parameters. For any given fine, however, one can distinguish cases analogous to those in Table 1 and obtain the corresponding implementability results.

8. DISCUSSION

In this paper, we contribute to the literature on delegated monitoring in which only the number of detections, and not the underlying monitoring intensity or the level of misbehavior, can be observed by the delegating authority (governor). This literature points to a problem of underenforcement in the sense that the first-best monitoring intensity need not be implementable. The reason is that when several monitoring intensities are observationally identical (that is, give rise to the same number of detections), the inspector to which the monitoring task is delegated can typically be induced to choose only the minimum of these. This also imposes a lower bound on the resulting crime level.

We first generalize the results from the previous literature and characterize the full set of implementable monitoring policies. We then perform a comparative-statics analysis to study how the set of implementable monitoring intensities varies with changes in the distribution of gains from the offense and in the fine. This set is small for low benefits from crime or high fines. In particular, the largest monitoring policy that is implementable decreases with the fine such that deterrence need not be increasing in the fine.

We then consider an extension of the model in which the utility of inspectors is directly influenced by the number of offenses in the population. In this respect, we believe that the intrinsic motivation to keep the level of crime low is an empirically relevant (and commonly neglected) factor. We first show that low levels of intrinsic motivation do not qualitatively affect the findings from the baseline model, and still only rel-

atively low monitoring intensities can be implemented. However, when the degree of intrinsic motivation is sufficiently large, a mirror result prevails, and the set of implementable policies is bounded from below instead. Consequently, either the desired monitoring policy is large enough to be implementable or overenforcement occurs; that is, for any payment scheme, the inspector will choose a monitoring intensity that is larger than the one the governor wants to implement.

Our results also suggest that when inspectors are crime sensitive (for example, because of intrinsic motivation), incentive schemes such as bounties as often discussed in the related literature might be even more delicate instruments than previously thought. The reason is that it crucially depends on the inspector's degree of intrinsic motivation whether such extrinsic incentives tend to reinforce the monitoring incentives generated by intrinsic motivation or crowd them out.

Our analysis, hence, points to the potentially crucial role of the degree of intrinsic motivation in the context of delegated monitoring, which, however, will typically be unobservable for governors. Therefore, an interesting extension of the model would be a screening framework in which the governor can offer different pairs of monitoring intensities $p(\hat{\beta})$ and payments $R(\hat{\beta})$, depending on the inspector's reported type $\hat{\beta}$ (which does not necessarily coincide with his true type β). In contrast to standard screening models, the resulting design problem for the governor becomes significantly more intricate because of additional incentive constraints. This is due to the fundamental property of this setting in which multiple monitoring intensities give rise to the same number of detections, which increases the scope of mimicking. A full analysis of such an extended screening framework is, however, beyond the scope of the present paper. From a more practical point of view, one possibility for learning about an inspector's intrinsic motivation over time would be to exploit the different comparative-statics properties of agents with different degrees of intrinsic motivation. For example, if it is possible to manipulate the costs of monitoring, then good types can be separated from bad types since the number of detections increases in only one of the two cases.

Finally, our analysis also sheds light on which contexts of delegated monitoring are more prone to issues such as under- or overenforcement and the ensuing consequences. In this respect, note that our framework applies not only to the arena of law and economics but in principle also to other settings of monitoring such as enforcing safety and ethical standards in the manufacturing industry or hygienic standards in the food

industry. Another relevant context is education, where a school authority delegates the task of educating pupils to schools and teachers. In doing so, it is an outsider in the sense that it can typically observe only which grades are awarded in a school. However, it does not observe whether, for instance, good grades are due to the fact that the school, its teachers, and the pupils are all hard working or whether they are the result of a (tacit) agreement among these parties to grade leniently. As a result, looking only at grades might not be very informative about the level of education of pupils or the quality of schools. In this context, however, by featuring state- or nationwide tests that all pupils must take, it seems that authorities have successfully implemented institutional changes to ameliorate the problem that schools of different quality might be observationally identical.²⁹ Even though such or similar measures might not always be available in the contexts to which our framework applies, our analysis nevertheless clearly points to the beneficial role of institutional changes that help authorities overcome their outsider status.

APPENDIX: PROOFS

A1. Proof of Theorem 1

For $\hat{p} \notin P_1$, there is a $\tilde{p} \in P_1$ such that $D(\tilde{p}) = D(\hat{p})$ by definition of P_1 . Noting that $R[D(\tilde{p})] = R[D(\hat{p})]$ while $C(\tilde{p}) < C(\hat{p})$, it follows that $u(\tilde{p}) > u(\hat{p})$, which shows that \hat{p} cannot be implemented. Now, suppose that $\hat{p} \in P_1$. Let $R(d) = C(p_1^d) + \varepsilon$ if $d = \hat{d} := D(\hat{p})$ and $R(d) = 0$ otherwise. Then $u(\hat{p}) = \varepsilon > 0$, while for small enough ε we have $u(p) \leq 0$ for all $p \neq \hat{p}$ because other monitoring intensities \tilde{p} that lead to the same number of detections (that is, $\tilde{p} \in P^{\hat{d}}$) are associated with higher costs, and all other choices (that is, $p \notin P^{\hat{d}}$) do not lead to any reward.

For the second statement, note that continuous G renders F continuous and thus renders D continuous as well. Since $D(p)$ starts with $D(0) = 0$ and reaches its global maximum for the first time at p^m , $D(p)$ attains every value of its image Δ in the interval $[0, p^m]$. Thus, for any $d \in \Delta$, $p_1^d \in [0, p^m]$.

A2. Proof of Proposition 1

Consider two distributions G and \tilde{G} , where G is first-order stochastically dominated by \tilde{G} ; that is, $G(g) \geq \tilde{G}(g)$ for all g . Denote by $\tilde{F}(p)$ and $\tilde{D}(p)$ the respective functions resulting under \tilde{G} .

i) For all $p \in [0, 1]$, we have

29. In Germany, for example, many states recently introduced mandatory statewide tests in German, English, and mathematics.

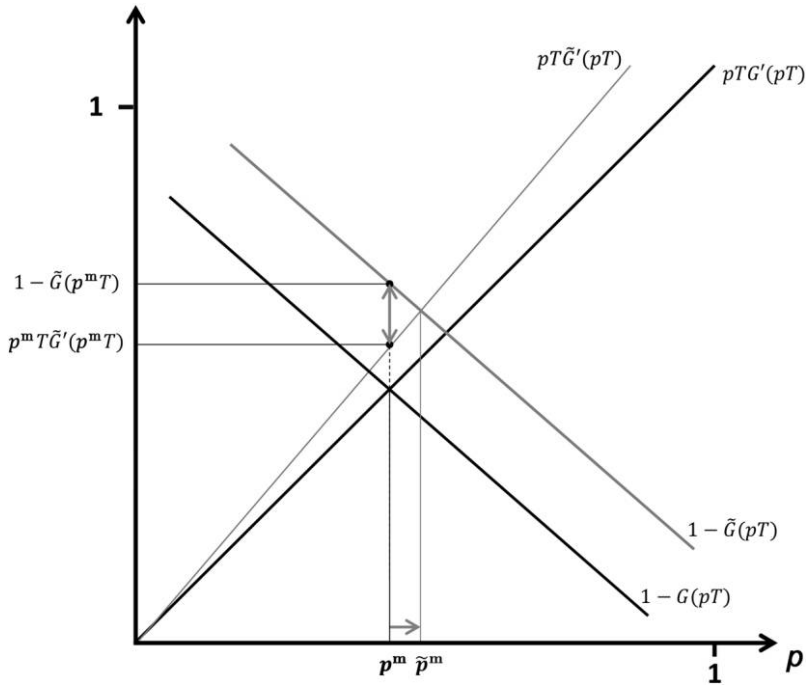


Figure A1. Marginal deterrence and detection effects for distributions G and \tilde{G}

$$F(p) = 1 - G(pT) \leq 1 - \tilde{G}(pT) = \tilde{F}(p)$$

and

$$D(p) = p[1 - G(pT)] \leq p[1 - \tilde{G}(pT)] = \tilde{D}(p).$$

ii) Under corollary 1.i, p^m and \tilde{p}^m are characterized by the first-order conditions $1 - G(pT) = pTG'(pT)$ and $1 - \tilde{G}(pT) = pT\tilde{G}'(pT)$. By corollary 1.ii, the right-hand sides of both equations are increasing in p . The left-hand sides of both equations are decreasing in p such that p^m and \tilde{p}^m are the only intersections. This is illustrated in Figure A1 (while the FOSD always implies that $1 - \tilde{G}(pT) \geq 1 - G(pT)$, the relation of the two right-hand sides is ambiguous). Observe that $\tilde{p}^m \geq p^m$ if and only if $1 - \tilde{G}(p^m T) \geq p^m T \tilde{G}'(p^m T)$. The assumption $\tilde{G}'(p^m T) - G'(p^m T) < [G(p^m T) - \tilde{G}(p^m T)]/(p^m T)$ is equivalent to

$$1 - \tilde{G}(p^m T) > p^m T \tilde{G}'(p^m T) + \underbrace{[1 - G(p^m T)] - p^m T G'(p^m T)}_{=0}, \quad (\text{A1})$$

which yields the result. Q.E.D.

A3. Proof of Proposition 2

To show the assertions for two fines T and \tilde{T} , we determine the slopes of the relevant functions.

i) For all $p \in (0, 1]$, we have

$$\frac{\partial F(p, T)}{\partial T} = -pG'(pT) < 0$$

and

$$\frac{\partial D(p, T)}{\partial T} = -p^2G'(pT) < 0.$$

ii) By corollary 1.i, $p^m(T)$ is interior. Using the implicit-function theorem, we have

$$\begin{aligned} \frac{\partial p^m(T)}{\partial T} &= - \left[\frac{\partial^2 D(p, T)}{\partial p \partial T} / \frac{\partial^2 D(p, T)}{\partial^2 p} \right] \\ &= - \frac{2p^m(T)G'[p^m(T)T] + p^m(T)^2TG''[p^m(T)T]}{2TG'[p^m(T)T] + p^m(T)T^2G''[p^m(T)T]}. \end{aligned}$$

By corollary 1.ii, $D(p, T)$ is concave, and thus the denominator is negative (this also follows from the second-order condition for $p^m(T)$ to maximize $D(p, T)$). Observe that the negative sign of the denominator is equivalent to $G''[p^m(T)T] < -2G'[p^m(T)T]/[p^m(T)T]$, which is also equivalent to the numerator being negative. Taken together, the expression is negative since the fraction after the minus sign is positive. Q.E.D.

A4. Proof of Proposition 3

For $C(p) = cp^2$ and $F(p) = 1 - p$, the inspector's optimization problem is $\max_p R[p(1 - p)] - \beta(1 - p) - cp^2$. Consider some desired monitoring intensity \hat{p} and let $\tilde{p} := 1 - \hat{p}$. Noting that $R[\hat{p}(1 - \hat{p})] = R[\tilde{p}(1 - \tilde{p})]$, we obtain $u(\hat{p}) \geq u(\tilde{p})$ if and only if $\beta(2\hat{p} - 1) \geq c(2\hat{p} - 1)$.

Suppose that $\beta > c$. Then $u(\hat{p}) > u(\tilde{p})$ if and only if $\hat{p} > \frac{1}{2}$ such that any desired intensity $\hat{p} < \frac{1}{2}$ is strictly worse than $\tilde{p} = 1 - \hat{p}$. Now, let $R(d) = x > \beta + c$ if $d = \hat{d} := D(\hat{p})$ and $R(d) = 0$ otherwise. Then, choosing x large enough, for example, $x = \beta + c$, yields $u(\hat{p}) = x - \beta(1 - \hat{p}) - c\hat{p}^2 > 0 - \beta(1 - p) - cp^2$ for any $p \neq \hat{p}, \tilde{p}$. The case $\beta < c$ is fully analogous. Q.E.D.

A5. Proof of Proposition 4

Suppose that the crime function is $F(p) = 1 - p$ and the inspector's utility is given by equation (3). If every detection is rewarded with some bounty r , the inspector's optimization problem becomes

$$\max_{p \in [0, 1]} rp(1 - p) - \beta(1 - p) - cp^2.$$

An interior solution is characterized by the first-order condition $p^*(r, \beta) = (r + \beta)/2(r + c)$, which is the first part of the proposition. The second part follows directly from

$$\frac{\partial p^*(r, \beta)}{\partial r} = \frac{c - \beta}{2(r + c)^2},$$

given, again, that the solution is interior.

An interior solution is obtained if $r \geq \max\{-c, -\beta, \beta - 2c\}$, where the first condition is equivalent to concavity of the maximization problem, while the second and the third conditions assure that the function is increasing at the boundary $p = 0$ and decreasing at the boundary $p = 1$. To show that $r \geq \max\{-\beta, \beta - 2c\}$ is sufficient for an interior solution $p^*(r, \beta)$ and for the sake of completeness, we formally provide the optimal monitoring policies for all possible types.

First, for $\beta < c$ (bad types), we have $-\beta > -c > \beta - 2c$. Thus, the optimal behavior is

$$p^*(r, \beta) = \begin{cases} \frac{r+\beta}{2(r+c)} & \text{if } r \geq -\beta \\ 0 & \text{if } r < -\beta \end{cases}.$$

For $r = -\beta$, the inspector chooses $p^* = 0$, and for growing r , $p^* = \frac{1}{2}$ is approached from the left.

Second, for $\beta > c$ (good types), we have $\beta - 2c > -c > -\beta$. Thus, the optimal behavior is

$$p^*(r, \beta) = \begin{cases} \frac{r+\beta}{2(r+c)} & \text{if } r \geq \beta - 2c \\ 1 & \text{if } r < \beta - 2c \end{cases}.$$

For $r = \beta - 2c$, the inspector chooses $p^* = 1$, and for growing r , $p^* = \frac{1}{2}$ is approached from the right.

Third, for the nongeneric case $\beta = c$, we have $-\beta = -c = \beta - 2c$. Then the optimal behavior is

$$p^*(r, \beta) = \begin{cases} \frac{r+\beta}{2(r+c)} = \frac{1}{2} & \text{if } r > -\beta \\ \{0, 1\} & \text{if } r < -\beta \\ [0, 1] & \text{if } r = -\beta \end{cases}.$$

For $r = -\beta$ the inspector is indifferent between all choices, for larger r he maximizes the number of detections, and for smaller r he minimizes them. Q.E.D.

A6. Proof of Proposition 5

We first establish that the monitoring intensity \hat{p} that maximizes equation (4) satisfies

$$\hat{p} \geq p^m = \frac{1}{2} \Leftrightarrow b \geq c - \beta + \frac{\pi}{2}. \quad (\text{A2})$$

If maximizing equation (4) yields an interior solution, then the condition above follows from a simple rearrangement of the first-order condition provided by equation (6). Otherwise, maximizing equation (4) either yields the boundary solution $\hat{p} = 0$ for $h + \beta < 0$, which implies that $h < c - \beta + \pi/2$, or yields the boundary solution $\hat{p} = 1$ for $h + \beta \geq 0$ and $c < (h + \beta - \pi)/2$, which implies that $h > 2c - \beta + \pi > c - \beta + \pi/2$. This shows the first assertion.

i) By proposition 3, $\beta < c$ implies that any p is implementable if and only if $p \in P_1 = [0, p^m]$. Hence, \hat{p} is implementable if and only if $\hat{p} \leq p^m$, which is satisfied if and only if $h \leq c - \beta + \pi/2$. Otherwise, we have $h > c - \beta + \pi/2$, which implies that $\hat{p} > p^m$, while only $p \leq p^m$ is implementable.

ii) By proposition 3, $\beta > c$ implies that any p is implementable if and only if $p \in [p^m, 1]$. Hence, \hat{p} is implementable if and only if $\hat{p} \geq p^m$, which is satisfied if and only if $h \geq c - \beta + \pi/2$. Otherwise, we have $h < c - \beta + \pi/2$, which implies that $\hat{p} < p^m$, while only $p \geq p^m$ is implementable. Q.E.D.

REFERENCES

- Andreoni, James. 1991. Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime? *RAND Journal of Economics* 22:385–95.
- Arlen, Jennifer. 1994. The Potentially Perverse Effects of Corporate Criminal Liability. *Journal of Legal Studies* 23:833–67.
- Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169–217.
- Becker, Gary S., and George J. Stigler. 1974. Law Enforcement, Malfeasance, and Compensation of Enforcers. *Journal of Legal Studies* 3:1–18.
- Besanko, David, and Daniel F. Spulber. 1989. Delegated Law Enforcement and Noncooperative Behavior. *Journal of Law, Economics, and Organization* 5:25–52.
- Bubb, Ryan, and Patrick L. Warren. 2014. Optimal Agency Bias and Regulatory Review. *Journal of Legal Studies* 43:95–135.
- Coşgel, Metin M., Haggay Etkes, and Thomas J. Miceli. 2011. Private Law Enforcement, Fine Sharing, and Tax Collection: Theory and Historical Evidence. *Journal of Economic Behavior and Organization* 80:546–52.
- Deci, Edward L. 1971. Effects of Externally Mediated Rewards on Intrinsic Motivation. *Journal of Personality and Social Psychology* 18:105–15.
- Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin* 125:627–68.
- Di Tella, Rafael, and Ernesto Schargrodsky. 2004. Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack. *American Economic Review* 94:115–33.

- Feess, Eberhard, and Ansgar Wohlschlegel. 2009. Why Higher Punishment May Reduce Deterrence. *Economics Letters* 104:69–71.
- Frey, Bruno S., and Reto Jegen. 2001. Motivation Crowding Theory. *Journal of Economic Surveys* 15:589–611.
- Frey, Bruno S., and Felix Oberholzer-Gee. 1997. The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out. *American Economic Review* 87:746–55.
- Garoupa, Nuno, and Daniel Klerman. 2002. Optimal Law Enforcement with a Rent-Seeking Government. *American Law and Economics Review* 4:116–40.
- Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. When and Why Incentives (Don't) Work to Modify Behavior. *Journal of Economic Perspectives* 21(4):1–21.
- Gneezy, Uri, and Aldo Rustichini. 2000a. A Fine Is a Price. *Journal of Legal Studies* 29:1–17.
- . 2000b. Pay Enough or Don't Pay at All. *Quarterly Journal of Economics* 115:791–810.
- Ichino, Andrea, and Gerd Muehlheusser. 2008. How Often Should You Open the Door? Optimal Monitoring to Screen Heterogeneous Agents. *Journal of Economic Behavior and Organization* 67:820–31.
- Kugler, Maurice, Thierry Verdier, and Yves Zenou. 2005. Organized Crime, Corruption, and Punishment. *Journal of Public Economics* 89:1639–63.
- Laffont, Jean-Jacques, and David Martimort. 2002. *The Theory of Incentives: The Principal-Agent Problem*. Princeton, NJ: Princeton University Press.
- Landes, William M., and Richard A. Posner. 1975. The Private Enforcement of Law. *Journal of Legal Studies* 4:1–46.
- Levitt, Steven D. 1997. Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *American Economic Review* 87:270–90.
- Levitt, Steven D., and Thomas J. Miles. 2007. Empirical Study of Criminal Punishment. Pp. 455–95 in vol. 1 of *Handbook of Law and Economics*, edited by A. Mitchell Polinsky and Steven Shavell. Amsterdam: North-Holland.
- Levitt, Steven D., and Sudhir Alladi Venkatesh. 2007. An Empirical Analysis of Street-Level Prostitution. Unpublished manuscript. University of Chicago, Department of Economics, Chicago.
- Nussim, Jacob, and Avraham D. Tabbach. 2009. Deterrence and Avoidance. *International Review of Law and Economics* 29:314–23.
- Polinsky, A. Mitchell. 1980. Private versus Public Enforcement of Fines. *Journal of Legal Studies* 9:105–27.
- Polinsky, A. Mitchell, and Steven Shavell. 2007. The Theory of Public Enforcement of Law. Pp. 403–54 in vol. 1 of *Handbook of Law and Economics*, edited by A. Mitchell Polinsky and Steven Shavell. Amsterdam: North-Holland.
- Stigler, George J. 1970. The Optimum Enforcement of Laws. *Journal of Political Economy* 78:526–36.

- Titmuss, Richard M. 1970. *The Gift Relationship: From Human Blood to Social Policy*. London: Allen & Unwin.
- Tsebelis, George. 1990. Penalty Has No Impact on Crime: A Game-Theoretic Analysis. *Rationality and Society* 2:255–86.